中华人民共和国国家标准

信息交换用汉字编码字符集 第二辅助集

UDC 681.3.043

GB 7589-87

Code of Chinese ideograms set for information interchange
The 2nd supplementary set

本标准规定了7237个汉字及其编码。这些汉字与GB 2312—80《信息交换用汉字编码字符集 基本集》一起,适用于汉字处理、汉字通信等系统之间的信息交换。

1 有关标准

- 1.1 GB 1988-80 《信息处理交换用的七位编码字符集》
- 1.2 GB 2311-80 《信息处理交换用七位编码字符集的扩充方法》
- 1.3 GB 2312-80 《信息交换用汉字编码字符集 基本集》

2 术语含义

- 2.1 基本集: GB 2312中规定的字符集。
- 2.2 辅助集:基本集以外的汉字编码字符集。
- 2.3 简化字: "繁体字"的对称。同一汉字,简化字比繁体字笔画为少。
- 2.4 繁体字: "简化字"的对称。原来笔画较多,汉字简化后已有简化字代替的字。
- 2.5 异体字: 音同、义同而形体不同的字。

3 图形字符

- **3.1** 图形字符代码表 图形字符代码表见表 1。
- **3.2** 空白位置 代码表中,凡未填入图形字符的空白位置,均作为进一步标准化区域。
- 3.3 图形字符的种类及数量 本标准收GB 2312 (基本集)以外的汉字7237个。

4 编码

4.1 七位代码系双字节编码表示

本标准在七位代码系中,对任意一个图形字符都采用两个字节表示,每个字节均采用GB 1988 及 GB 2311中的七位编码表示。两个字节中,位于前面的高位字节为第一字节,位于后面的低位字节为第二字节。其在代码表中的位置均分别为 $2/1 \sim 7/14$ (十进制)或 $21 \sim 7$ E (十六进制),见表 1。

在七位代码系中,指明本标准双字节编码表示图形字符集的转义序列为: ESC 2/4 2/8 F,ESC 2/4 2/9 F,ESC 2/4 2/10 F 及ESC 2/4 2/11 F。

注:终止字符户的值有待确定。

4.2 八位代码系双字节编码表示

本标准在八位代码系中,对任意一个图形字符都采用两个字节表示,每个字节均采用GB 2311中的八位编码表示。两个字节中,位于前面的高位字节为第一字节,其在八位代码表中的位置为 10/01 ~ 15/14 (上进制)或A 1 ~ F E (十六进制);位于后面的低位字节为第二字节,其在八位代码表中的位置为02/01 ~ 07/14 (十进制)或21 ~ 7 E (十六进制),见表 1 。

4.3 图形字符在代码表中位置的表示方法

代码表纵向分成94个区,每个区横向分成94个位。区的编号为1~94,由高位字节标识;位的编号也为1~94,由低位字节标识。代码表中的任何一个图形字符的位置用它所在的区号与位号标识。区号与位号之间用连字符相连。

例 汉字"丏"用16-1表示。

注:① 区号和位号与GB 1988中的列号与行号的对应关系为:

列号/行号2 / 1 ~ 7 / 14区号1 ~ 94位号1 ~ 94

② 区号和位号与GB 2311八位代码表中列号和行号的对应关系为:

列号/行号 10/01~15/14 对应区号 1~94 列号/行号 02/01~07/14 对应位号 1~94

汉字分区表

附录 A 汉字的选择和排列 (参考件)

A.1 汉字的选择

本标准所选汉字是GB 2312 (基本集)未收的汉字,它们是汉语通用的规范汉字,它们的用途与使用率一般低于基本集中的汉字,但是高于第四辅助集中的汉字。

这些汉字是以《新华字典》、《现代汉语词典》、《辞海》为基础,参考《标准电码本》及《北京图书馆检字表》,同时参考对《汉语大字典》、《康熙字典》、《中华大字典》及《中文大辞典》全部汉字的义项数统计资料和最近发表的有关汉字频度统计结果以及国内大型汉字系统所用的汉字,包括科技专业用字,共选收了7237个汉字。

本标准所收汉字的字形符合中华人民共和国文化部、中国文字改革委员会公布的《印刷通用汉字字形表》与中国文字改革委员会公布的《简化字总表》的规定。

本标准以简化字为规范字体,原则上不收繁体字和《第一批异体字整理表》中被淘汰的异体字。除少数汉字外,类推简化的原则按《简化字总表》中第二表规定。

A.2 汉字的排列

本标准中7237个汉字的排列按照1983年中国文字改革委员会与文化部出版局等单位联合制订的统一汉字部首排检法草案规定的201个部排列。部首次序按画数排列,同部首字按部首以外的画数排列,同画数的字以笔形顺序横(一)、直(一)、撇(丿)、点(、)、折(乙,包括一、门、 し等笔形)为序,起笔相同按第二笔,依次类推。附汉字分区表。

47区 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 01 1

55 区 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19 19 19 社种设防柘枯祔括桧榆械袯祾祾植裺裸褐禖 20 建褐穗稀镇襜襜穑槽禤禫襚獭獭å犨虫矼 59 社矻砐似双矸砺砋砜张砎枞砼砼砏砃砊砄砞矻 59 菪砵呵砳砠砍珍砱鸹砍砫岩砮岩降径砯䂮硈碧 60 79 替砵呵砳砠砍珍砱鸹称砫岩砮岩降径砯䂮硈碧

70 🔀 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19

71 🗵 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19

88 🔀 01 02 03 04 05 06 07 08 09 10 11 12 13 14 15 16 17 18 19

附 录 B 编码方式及使用 (参考件)

B.1 编码范围

信息交换用汉字编码字符集将逐步对汉字、各少数民族文字及其它常用图形字符进行编码。

B.2 七位代码系双字节编码表示

根据GB 1988及GB 2311,七位代码系双字节编码表示就是每个图形字符都采用两个GB 1988及GB 2311中的七位编码表示,并仅用 $2/1 \sim 7/14$ 共94个码位。这样,一个双字节编码字符集纵向由高位字节分成94个区,横向由低位字节分成94个位,最多可以编码94×94 = 8836个图形字符。

当需要编码的图形字符数超过8836时,根据GB 2311可以采用多个双字节编码字符集。每个字符集分别采用转义序列与调用字符进行相应的指明和调用(详细使用参见GB 2311)。

各集汉字从16区开始连续排列,1~15区作为保留标准区域。

B.3 八位代码系双字节编码表示

如上所述,采用七位代码系双字节编码表示,一个双字节编码图形字符集只能编码8836个图形字符,超过8836时,需要采用转义序列及调用字符。这在某些场合处理不太方便,为此,本标准规定也可采用八位代码系双字节编码表示,即每个汉字用两个字节,每个字节用八位位组表示,其中每个位组都采用GB 2311八位代码系中的八位代码表示。并且,为避开GB 2311八位代码表中 C0 集与 C1 集的控制功能,两个八位位组分别只使用位置02/01~07/14(十六进制:21~7 E)及位置10/01~15/14 (十六进制:A 0~F F)的代码值。这样,两个八位位组构成的双字节图形字符集最少可编码(94 + 94)×(94 + 94)= 4×(94×94)= 35344个汉字或图形字符。

B.4 简化字与繁体字分别对应编码

根据我国简化字与繁体字并用的实际情况,而简化汉字又是我国法定的规范字形,为便于汉字信息的处理与交换,简化字与繁体字分别对应编码。即,原则上将"规范字集"内的简化字改成相应的繁体字就成为"繁体字集"。规范字集的序号用偶数表示;繁体字集的序号用奇数表示。这样,基本集、第二辅助集、第四辅助集、……为规范字集;第一辅助集、第三辅助集、第五辅助集、……为对应的繁体字集。

B.5 图形字符集之间的转换

单字节图形字符集与本标准八位代码系双字节图形字符集之间以及七位代码系双字节图形字符集与本标准八位代码系双字节图形字符集之间转换的国际标准转义序列或控制功能目前还未规定。为此,使用时须事先约定。

附加说明:

本标准由中华人民共和国电子工业部提出。

本标准由电子工业部华北计算技术研究所、国家语言文字工作委员会汉字处负责起草。

本标准主要起草人陈耀星、傅永和、高景成、魏励。